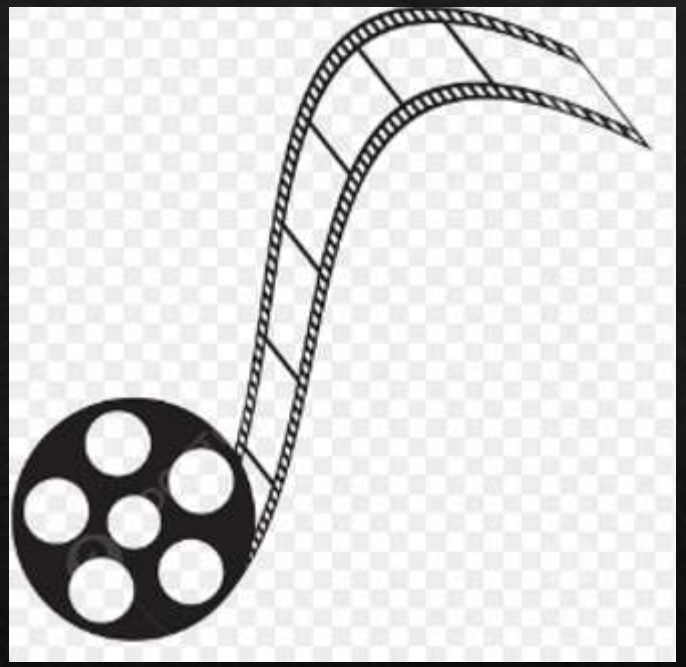
ECHOREEL:
A CONTENT-BASED MOVIE RECOMMENDATION SYSTEM 🎬



By: Sneha Kumari
(IITRPRAI_**24081716**)

Course: Minor in Artificial Intelligence, IIT Ropar

Track: Artificial Intelligence

INTRODUCTION

- In today's world of streaming platforms, viewers face **content overload**.
- It becomes challenging to decide **what to watch next**.
- EchoReel aims to make this easier by using **Machine Learning** to recommend movies **similar to your favorites**.
- This system learns **movie features and similarities** — not popularity — to generate meaningful suggestions.

🌍 THE MODERN ENTERTAINMENT DILEMMA

- In today's streaming-driven era, users face an overwhelming number of choices — thousands of movies, limited time, and constant indecision.
- **Key Challenges:**
- 🎥 10,000+ movies released globally each year.
  🔎 Users spend an average of 20+ minutes daily just deciding what to watch.
  💡 OTT fatigue is real — personalization is now a necessity, not a luxury.
- **EchoReel** addresses this challenge by offering intelligent, data-driven movie recommendations that *truly align* with the user's taste.

THE AI REVOLUTION IN MOVIE DISCOVERY

- Movie discovery has evolved from simple genre filters to context-aware AI models. EchoReel represents the next step — combining machine learning and NLP for story-level understanding.
- **Evolution Timeline:**

Rule-Based Filters: Basic genre/year/actor sorting.
Collaborative Filtering: "Users who liked X also liked Y."
Content-Based Filtering (EchoReel): Analyzes movie plots, cast, and crew to find meaningful similarities.
Deep Learning Future: Transformer-based embeddings (e.g., BERT) for semantic understanding.

🔍 HOW ECHOREEL WORKS

- EchoReel uses **Natural Language Processing (NLP)** and **Cosine Similarity** to understand each movie's narrative DNA and recommend contextually similar titles.
- **Pipeline Overview:**

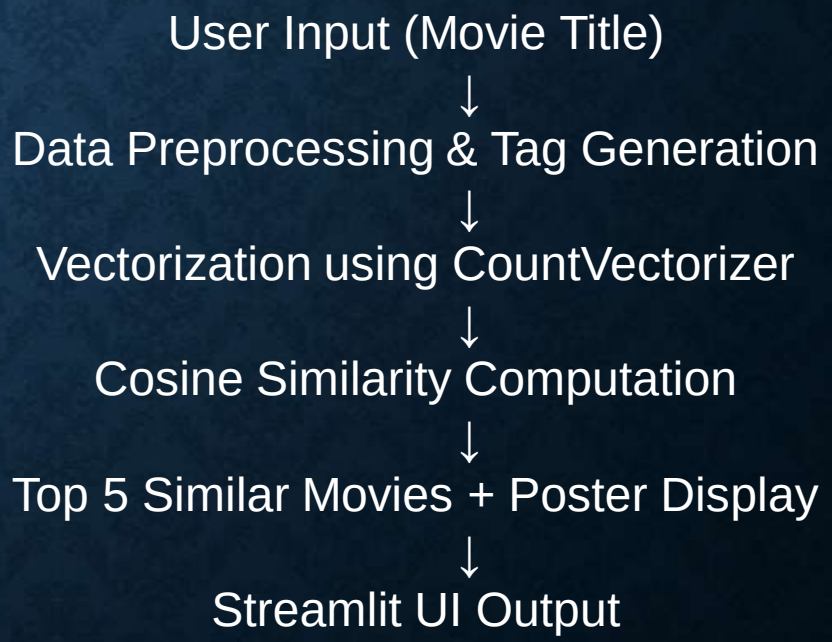
Data Source: Movies dataset fetched from **Kaggle**: ML-Based-Movies-Recommender/dataset/tmdb_5000_credits.csv and tmdb_5000_movies.csv
Preprocessing: Extracts *overview*, *genres*, *cast*, *keywords*, and *crew* data
Feature Engineering: Merges all text features into a unified "tags" column
Vectorization: Converts text data into numerical form using **CountVectorizer**
Similarity Computation: Uses **Cosine Similarity** to find the closest movies in vector space
Poster Retrieval & Output: Top 5–10 recommendations displayed via **Streamlit**, with posters fetched using the **OMDB API**

⚙️ TECHNOLOGY STACK

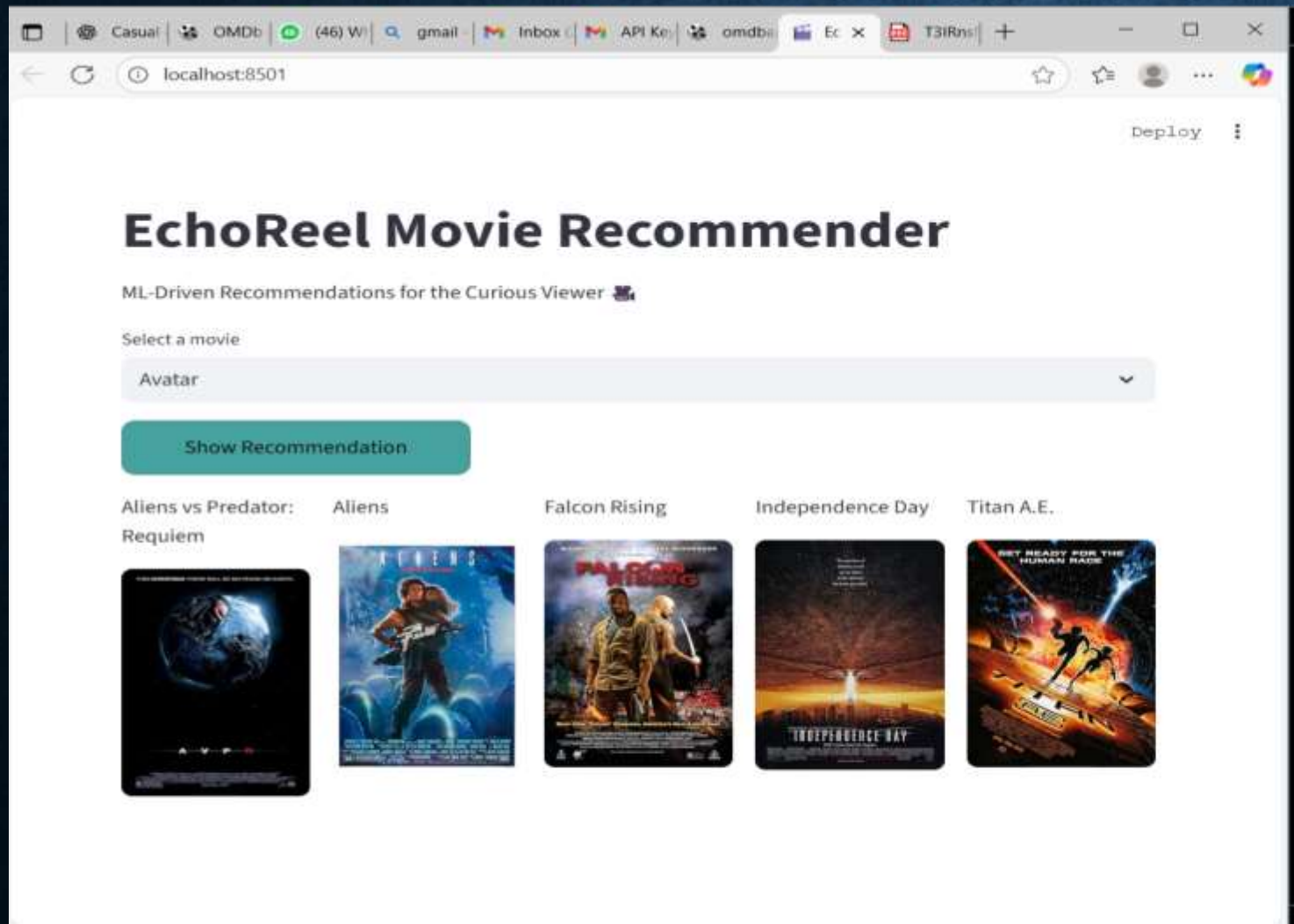
Component	Technology Used
Programming Language	Python
Libraries	Pandas, NumPy, Scikit-learn, NLTK
Dataset	Kaggle: tmdb_5000_credits.csv & tmdb_5000_movies.csv
API	OMDB API (for movie posters)
ML Techniques	CountVectorizer, Cosine Similarity
Deployment	Streamlit
Model Saving	Pickle serialization

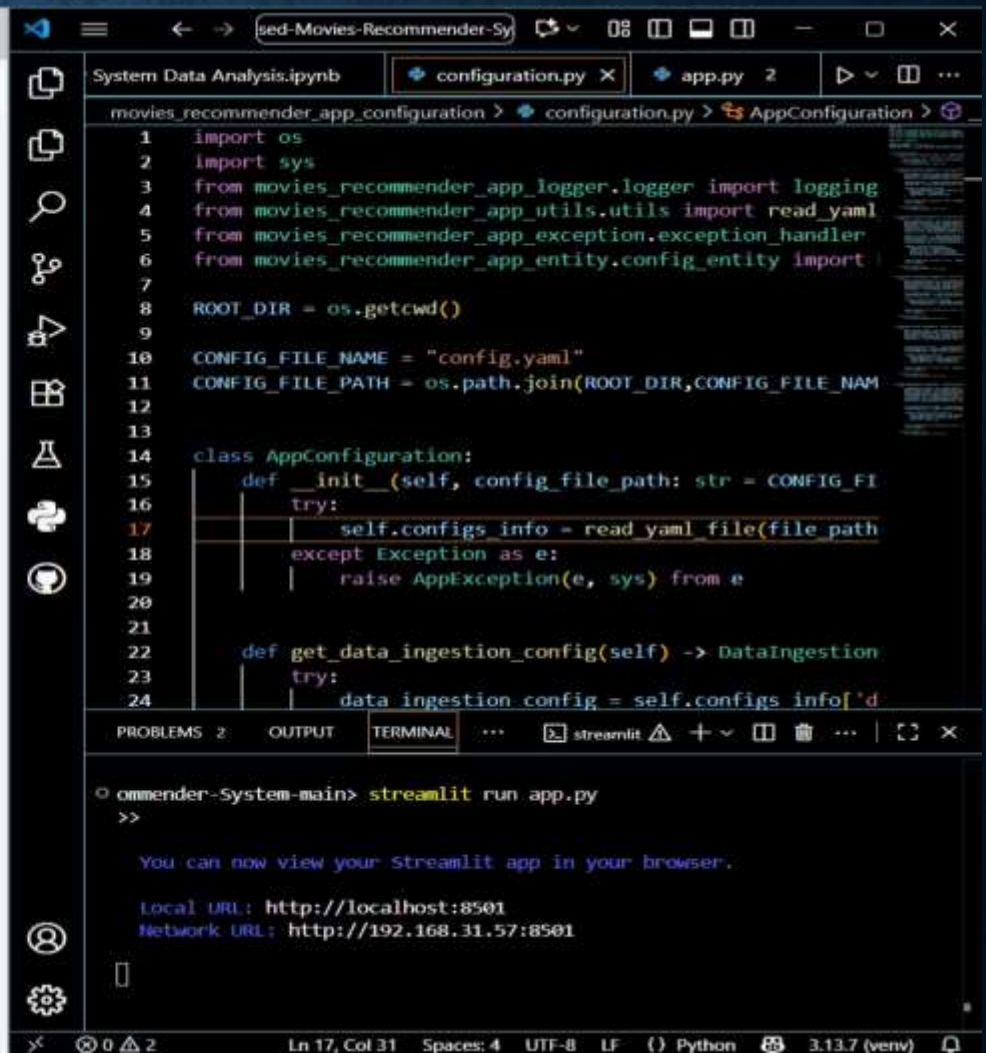
IMPLEMENTATION PROCESS AND SYSTEM ARCHITECTURE

- **Stage 1:** Data Ingestion → Loading and merging CSVs
- **Stage 2:** Data Validation → Removing null & duplicate entries
- **Stage 3:** Data Transformation → Generating tags, vectorization
- **Stage 4:** Model Training → Computing cosine similarity matrix
- **Stage 5:** Deployment → Building and running Streamlit app



APPLICATION INTERFACE (DEMO)



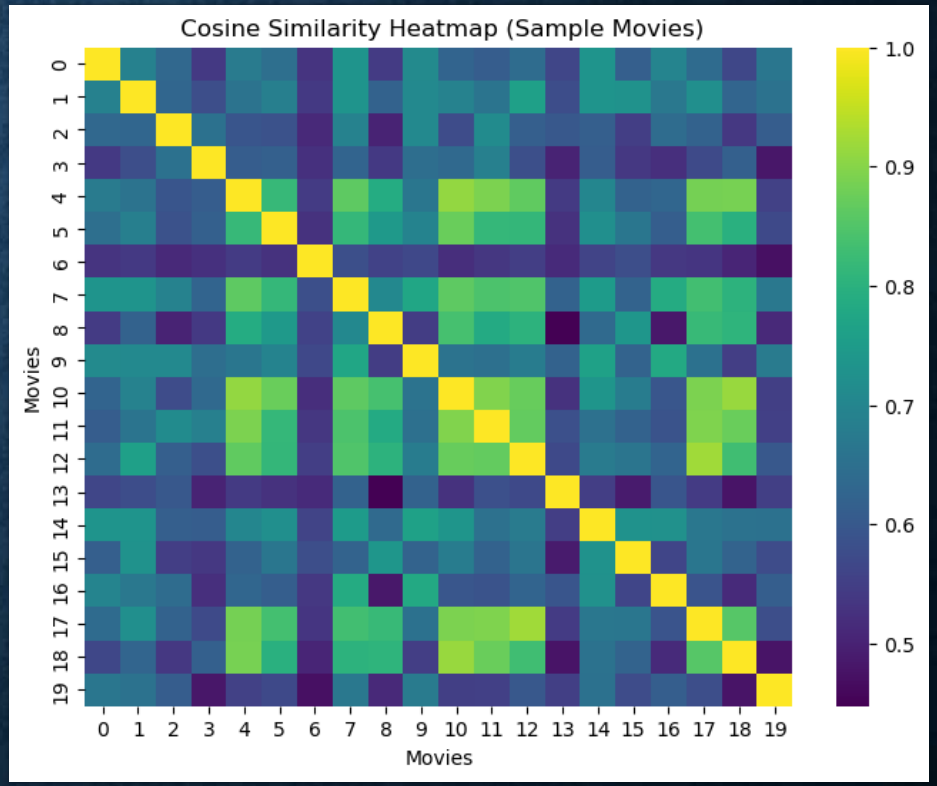


🎯 EVALUATION METRICS

Metric	Purpose	Result
Precision@10	Measures relevance among top 10 recommendations	0.50
Recall	Fraction of relevant movies retrieved	1.00
F1-Score	Balances precision and recall	0.67
Cosine Similarity Score	Measures vector-level closeness between movies	Most values 0.6–1.0; high intra-movie similarity on diagonal; clusters indicate thematic/genre grouping

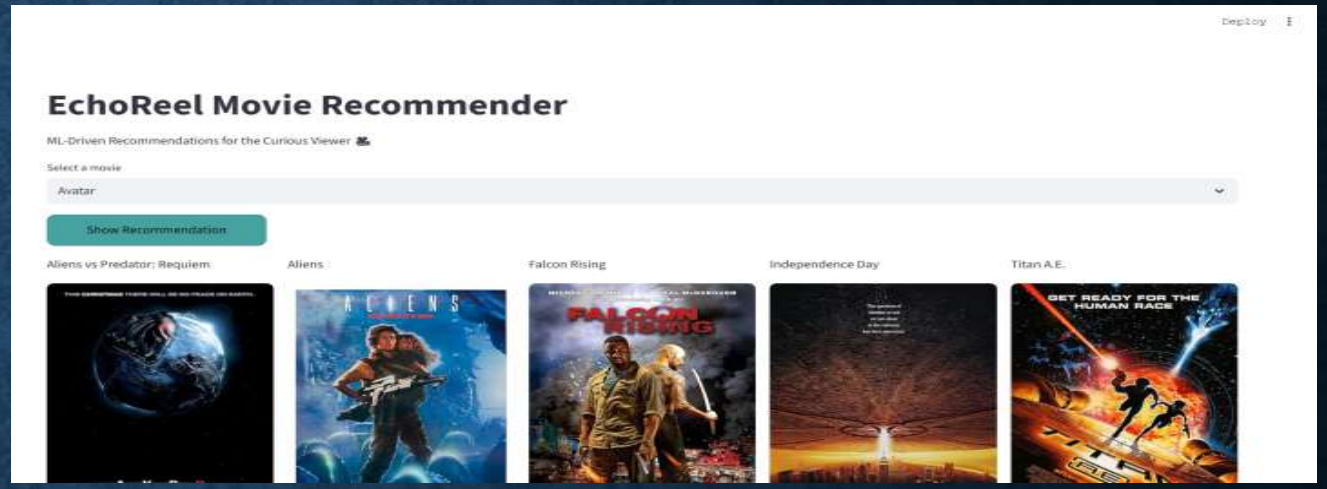
HEATMAP OF SIMILARITY SCORES

High intra-movie similarity: Diagonal shows perfect self-similarity
Clusters of similar movies: Bright patches indicate thematic/genre-based grouping
Low similarity zones: Dark areas indicate distinct content/features
Range-bound similarity: Most values between 0.6 and 1.0, indicating moderate-to-strong relationships



🏆 RESULTS & OBSERVATIONS

- **User Input:** 🎞 *Avatar (2009)*
  EchoReel Output:
  ➞ Aliens vs Predator: Reqiem
  ➞ *Falcon Rising*
  ➞ Aliens
  ➞ *Independence Day*
  ➞ Titan A.E.



- ✨ Each recommendation shares **sci-fi, action, and futuristic storytelling themes**, showing how EchoReel captures narrative and stylistic context.

✅ Fast similarity computation due to precomputed vectors
✅ Consistently relevant recommendations across genres
✅ Real-time poster fetching enhances engagement
✅ Streamlit UI provides clean and intuitive experience

INDUSTRY IMPACT

- AI-powered recommendation systems are reshaping entertainment worldwide:
- **OTT Platforms:** 80% of Netflix views originate from AI recommendations
  User Engagement: Personalized AI suggestions increase retention by 35%
  Market Growth: Global recommendation engine market projected to reach **$17B by 2028**
  Scalability: EchoReel's OMDB-based model can adapt to multiple languages and content ecosystems

ETHICS & DATA RESPONSIBILITY

- EchoReel emphasizes ethical, transparent, and privacy-respecting AI design:
- No personal data collection — content-based only
- Avoids user bias by not depending on ratings or demographics
- Transparent similarity logic and explainable recommendations

🌱 SUSTAINABILITY IN AI SYSTEMS

- EchoReel optimizes for lightweight, sustainable computation:
- Uses **CountVectorizer** over heavy neural models for speed and energy efficiency
- Cached similarity matrices reduce redundant computation
- Compact Pickle model = lower memory footprint and greener AI processing

🚀 THE ROAD AHEAD

- EchoReel is just the beginning of smarter, context-aware entertainment discovery.
- **Future Enhancements:**
- 🔹 Integrate hybrid filtering (content + user behavior)
  🔹 Use **BERT embeddings** for semantic-level movie understanding
  🔹 Add visual similarity (poster-based CNN models)
  🔹 Deploy on cloud (AWS / Hugging Face Spaces) for global access

💡 CONCLUSION: EMPOWERING STORIES THROUGH AI

- EchoReel bridges technology and storytelling — helping users find the *right movie at the right time*.
- By combining **AI precision** with **human curiosity**, it redefines entertainment discovery for the modern era.
- 🎬 *"Because sometimes, the best story is the one you didn't know you needed to see."*

THANKYOU